(The amendments do not add new matter to the specification. Claims are being cancelled.)

The applicants hereby request the following amendments to the application upon entry into the U.S. National Stage:

Amendments to the background of the application:

1) page 5 line 22 after "territory" insert --and that it was difficult to predict the power of using a less dense map at that time--

2) page 5 line 22 after "10^{n}" insert -- **The inventor's work, however, is a predictor of the power and success of a less dense map.**--

A replacement page 5 with header "PCT/US99/04376(U.S. National Stage Entry Aug. 2000)" is enclosed to effect these amendments 1) and 2) to the background.

3) page 6 line 22 after "TDT," change the text "***to increase the likelihood of conditions occurring that increase the power of the TDT in the linkage study, the bi-allelic markers used in the study are chosen so that the least common allele frequencies of the markers vary systematically over a range or subrange of least common allele frequency.***" from bold face italics to regular italics with underlining. A replacement page 6 with header "PCT/US99/04376(U.S. National Stage Entry Aug. 2000)" is enclosed to effect amendment 3) to the background.

4) page 7 line 6 after "TDT," change the text "***to increase the likelihood of both criteria (1) and (2) occurring for one or more markers, so as to increase the power of the TDT in the linkage study, the bi-allelic markers used in the study are chosen so that the least common allele frequencies of the markers vary systematically over a range or subrange of least common allele frequency AND the chromosomal location of the markers vary systematically over one or more chromosomes or chromosomal regions. And the bi-allelic markers are chosen so that the markers' chromosomal locations and least common allele frequencies vary systematically in an essentially independent manner.***" from bold face italics to regular italics with underlining.

5) page 7 line 32 delete the text in brackets [***In addition, the two-dimensional linkage study techniques do not necessarily favor using markers in a scan that are about evenly spaced along a chromosome as in the conventional techniques.*** *This is because*]. On page 7 line 31 after the text "unfavorably" insert on the insert --**Conventional techniques use a one-dimensional concept of "closeness".** ***These techniques space markers about evenly along a chromosome in the hope that some markers will be "close" (on the chromosome) to the sought gene. (They also favor bi-allelic markers with least common allele frequencies near 0.5.) These***--.

A replacement page 7 with header "PCT/US99/04376(U.S. National Stage Entry Aug. 2000)" is enclosed to effect amendments 4) and 5) to the background.

6) page 8 line 18 insert on the next line after "background" -- **Summary**

Versions of the invention use a new, two-dimensional concept of "closeness" for association-based linkage studies. Versions of the invention use bi-allelic markers that "cover" or are distributed approximately evenly (or systematically) over two-dimensional regions. These regions have the two dimensions of chromosomal location and least common allele frequency. **Conventional techniques suffer from a kind of one-dimensional lack of depth perception.** (They also favor bi-allelic markers with least common allele frequencies near 0.5.) Two-dimensional linkage study techniques overcome this lack of depth perception. These two-dimensional techniques greatly increase the chance that one or more markers used in a study will be close to the sought gene in two-dimensions. This results in more powerful, systematic and efficient methods (including computer programs) and machines for finding genes, such as harmful

genes and genes of only modest effect. These techniques also use less dense (more efficient) marker maps (or marker "coverings").

The basic principles behind the two-dimensional approach spawn numerous other inventions. These include methods, machines and compositions of matter (groups of molecules) used for gathering the data (i.e. genotype/sample allele frequency data) used in the new two-dimensional studies, and computer techniques for using and handling such data. These techniques work for creatures other than human beings. And they work for markers and genes that are not bi-allelic (any marker or gene can be mathematically transformed to behave like it is bi-allelic). This summary is not exhaustive or limiting, there are other inventions not listed or specifically described here.--

A replacement page 8 with header "PCT/US99/04376(U.S. National Stage Entry Aug. 2000)" is enclosed to effect amendment 6) to the background.

Amendments to the Description

7) page 38 line 2, page 38 line 17 and page 38 line line 20 delete the text "Best Mode" and replace the text "Best Mode" with the text "Set/Subset Example". A replacement page 38 with header "PCT/US99/04376(U.S. National Stage Entry Aug. 2000)" is enclosed to effect the amendments to the description under item 7).

8) page 43 line 4 delete the text "Best Mode" and replace the text "Best Mode" with the text "Set/Subset Example". A replacement page 43 with header "PCT/US99/04376(U.S. National Stage Entry Aug. 2000)" is enclosed to effect the amendment to the description under item 8).

9) page 46 line 24 and line 28 delete the text "Best Mode" and replace the text "Best Mode" with the text "Set/Subset Example". One page 46 lines 26 and 27 delete the text "Best Mode" and replace the text "Best Mode" with the text "Set/Subset". A replacement page 46 with header "PCT/US99/04376(U.S. National Stage Entry Aug. 2000)" is enclosed to effect the amendments to the description under item 9).

Canceling of Claims and presentation of uncancelled claims for examination

The applicants hereby request that all claims in the application be cancelled except for the following claims that were filed April 17, 2000: Claims 3, 4, 5, 7, 8, 20, 21, 22, 23, 33, 34, 35, 37, 38, 50, 51, 52, 53, 54, 57. Thus the applicants request that only claims 3, 4, 5, 7, 8, 20, 21, 22, 23, 33, 34, 35, 37, 38, 50, 51, 52, 53, 54, 57 filed April 17 2000 be examined.

I hereby attest that no new matter is added to the specification of the application by the amendments requested in the two pages of this letter.

Respectfully submitted,

Robert McGinnis

Robert McGinnis
U.S. Patent Agent 44, 232

0.5/0.5. Secondly, bi-allelic markers with lower least common allele frequencies, less than 0.3(0.7/0.3) or 0.2(0.8/0.2), are viewed unfavorably for linkage studies in this reference. Thirdly, the early version of the criterion of "information content" of markers used in this reference was based on sib pair analysis and the later, current version of the criterion, does not depend on any particular test for linkage.[5, 6] **Thus, the criterion of information content in this reference, has never specifically employed th TDT (transmission disequilibrium test) or any association based test, whereas the two-dimensional linkage study techniques of this application are based on a completely different perspective of using association based tests.** (This reference[4] is not admitted to be prior art with respect to the present invention by it's mention in this background.)

Increased Power of the TDT (transmission disequilibrium test)

Characteristics of a new type of linkage test, the TDT (transmission disequilibrium test), were described in 1993. The inventor, R.E.McGinnis, was one of the authors of this reference.[7] In 1996, Risch and Merikangas argued that conventional linkage analysis has limited power to detect genes of modest effect. And Risch and Merikangas attempted to illustrate the increased power of association based linkage tests such as the TDT over other types of conventional linkage tests.[8] However, Risch and Merikangas' analysis was criticized by Muller-Myhsok and Abel as being based on the optimal assumption that the analyzed allele was the disease allele itself. Muller-Myhsok and Abel concluded that researchers should be aware that the power of association studies such as the TDT can be greatly diminished in more common, less optimal situations.[9] In their response to Muller-Myshok and Abels' letter, Risch and Merikangas essentially agreed with the logic of Muller-Myshok and Abels' criticism. Risch and Merikangas stated that to a large extent, the expectation with respect to linkage disequilibrium across the genome is uncharted territory and that it was difficult to predict the power of using a less dense map at that time.[10] **The inventor's work, however, is a predictor of the power and success of a less dense map.** (None of the references in this paragraph [7,8, 9,10] is admitted to being prior art with respect to the present invention by their mention in this background.)

More Detailed Studies of the Power of the TDT

The inventor, R.E.McGinnis, has done extensive investigations on the power of the TDT. His observations and calculations of the increased power of the TDT in many situations have been

---

[5] Kruglyak, et. al.: Complete Multipoint Sib-Pair Analysis of Qualitative and Quantitative Traits. Am J Hum Genet, 1995, vol. 57: pp. 439-454.

[6] Kruglyak, et. al.: Parametric and Nonparametric Linkage Analysis: A Unified Multipoint Approach. Am J Hum Genet , 1996, vol. 58, pp. 1347- 1363.

[7] Spielman, R.S., McGinnis, R.E., Ewens, W.J.: Transmission Test for Linkage Disequilibrium: The Insulin Gene Region and Insulin-dependent Diabetes Mellitus(IDDM). Am J Hum Genet, 1993, vol. 52, pp. 506-516.

[8] Risch, N. and Merikangas, K.: The Future of Genetic Studies of Complex Human Diseases. Science, 13 September 1996, vol. 273, pp. 1516-1517.

[9] Muller-Myshok, B. and Abel, L.: Technical Comments: The Future of Complex Diseases. Science, 28 February 1997, vol. 275, pp. 1328-1329.

[10] Risch, N. and Merikangas, K.: Technical Comments: The Future of Complex Diseases. Science, 28 February 1997, vol. 275, p. 1330.

published.[11] In this paper a general framework for determining the power of the TDT in many different situations is presented. The analysis of Risch and Merikangas [8] and others is shown by the inventor to be a special case of his general framework. His observations and calculations published in this paper have shown that the TDT has increased power in more common, less optimal situations as well as the less common, optimal situation cited by Muller-Myshok and Abel [9]. As opposed to the observation of Muller-Myhsok and Abel, the inventor's calculations indicate that association tests such as the TDT have increased power in typical situations even when the ratio m/p departs significantly from unity and, or the linkage disequilibrium between the analyzed (marker) allele and disease polymorphism is only half its maximum possible value. The inventor arrived at these conclusions independently and did not derive them from others.

A Major Conclusion Drawn by the Inventor about the TDT and Linkage Studies: Using Bi-allelic Markers of Systematically Varying Allele Frequencies Increases the Power of Linkage Studies Using the TDT

The inventor's calculations and observations about the increased power of the TDT in more common, less optimal situations led him to the conclusion that the power of linkage studies using the TDT is greatly increased under some conditions. Under some conditions, the power of the TDT in a linkage study using bi-allelic markers is greatly increased when each of one or more of the bi-allelic markers used in the study fulfill two criteria: (1) the allele frequencies of each of the one or more of the bi-allelic markers are similar (but not necessarily the same, or even approximately the same) as the allele frequencies of an unknown bi-allelic gene causing a disease under study; and (2) each of the one or more bi-allelic markers is in some degree of linkage disequilibrium with the gene. Thus for a typical linkage study using bi-allelic markers and the TDT, *to increase the likelihood of conditions occurring that increase the power of the TDT in the linkage study, the bi-allelic markers used in the study are chosen so that the least common allele frequencies of the markers vary systematically over a range or subrange of least common allele frequency.* This major conclusion of the inventor's research is quoted directly from his unpublished manuscript that was included with previously filed U.S. Provisional Patent Applications:"This example is typical and highlights perhaps the most important finding of this paper; namely the importance of using bi-allelic markers with heterozygosity similar to that of a bi-allelic disease gene. Indeed, since a majority of susceptibility loci may be bi-allelic, the judicious use of bi-allelic markers of both high, medium and low heterozygosity may be crucial in order to detect and replicate linkages to loci conferring modest disease risk." (page 25) (In this context the phrase "bi-allelic markers with heterozygosity similar to that of a bi-allelic disease gene" is essentially equivalent to "bi-allelic markers with individual allele frequencies similar to those of a bi-allelic disease gene" and "bi-allelic markers of both high, medium and low heterozygosity " is essentially equivalent to the phrase "bi-allelic markers whose least common individual allele frequencies are high, medium and low".)

Systematically Varying Both Marker Chromosomal Location and Marker Allele Frequency of Markers in Linkage Studies

[11] McGinnis, R.E.: Hidden Linkage: Comparison of the affected sib pair (ASP) test and transmission disequilibrium test (TDT). Annals of Human Genetics, 1998, vol. 62, pp. 159-179.

The inventor's calculations and observations have demonstrated the increased power of the TDT in more common, less optimal situations when a bi-allelic marker and bi-allelic gene have (1) similar but not identical allele frequencies and (2) the marker and gene are in some degree of linkage disequilibrium. Thus, for a typical linkage study using bi- allelic markers and the TDT, *to increase the likelihood of both criteria (1) and (2) occurring for one or more markers, so as to increase the power of the TDT in the linkage study, the bi-allelic markers used in the study are chosen so that the least common allele frequencies of the markers vary systematically over a range or subrange of least common allele frequency AND the chromosomal location of the markers vary systematically over one or more chromosomes or chromosomal regions. And the bi-allelic markers are chosen so that the markers' chromosomal locations and least common allele frequencies vary systematically in an essentially independent manner.*

Two-dimensional Linkage Study Techniques

As has been stated, conventional linkage study scanning techniques use markers that are distributed approximately evenly in the dimension of chromosomal location. These conventional, one dimensional, scanning techniques focus primarily on the chromosomal location of markers used in a scan and give little attention to the dimension of allele frequency.[1, 2, 3]

One of the main implications of the inventor's work is to use a set of bi-allelic markers for a typical linkage study using the TDT (or other association-based linkage test) wherein the chromosomal locations and least common allele frequencies of the markers in the set systematically vary in an essentially independent manner over the dimensions of chromosomal location and least common allele frequency respectively. This is equivalent to using a set of bi-allelic markers for a linkage study scan wherein the set of markers systematically scan or "cover" a two-dimensional region having dimensions of chromosomal location and least common allele frequency. (Such a two-dimensional region can be thought of as an area in an x-y plot or a group of squares on a chessboard.)

In addition, the inventor's calculations and observations indicate that bi-allelic markers having least common allele frequencies less than 0.3, 0.2 or even less than 0.1 have an important place in linkage studies using association based linkage tests. This is markedly different than Kruglyak's information content evaluation of bi-allelic markers for use in linkage studies, in which bi-allelic markers with least common allele frequencies less than 0.3 or 0.2 are viewed unfavorably.[4]

Conventional techniques use a one-dimensional concept of "closeness". ***These techniques space markers about evenly along a chromosome in the hope that some markers will be "close" (on the chromosome) to the sought gene. (They also favor bi-allelic markers with least common allele frequencies near 0.5.) These conventional techniques suffer from a kind of one dimensional view or lack of depth perception.*** **In the conventional techniques, a marker can look very close to a gene's location in terms of chromosomal location, but the marker can be very far from the gene's location in the new two-dimensional view used by versions of the invention. It is as if the conventi nal 1D techniqu s l ok at a chessboard fr m on dge. Markers and a gene which are on different squares of the b ard, but in the same column of squares, lo k v ry close to each other when the board is l oked at from on edg . But when the b ard is looked at**

from the top in 2D, tw dim nsions, markers which looked very close to ach other and th gene before (when looking from on edge) can be seen to be very far from the gene.

Further Implications of the Two-dimensional Linkage Study Perspective

These two-dimensional techniques work when multiple genes cause a genetic characteristic and are effective in searching for these genes. A two-dimensional bi-allelic marker "covering" or scanning approach also increases the power of linkage studies using other association based linkage tests such as the AFBACmethod, the haplotype relative risk (HRR) method [12], and comparison of marker allele frequencies in disease cases and unrelated controls[13]. These references[12, 13] are not admitted to being prior art with respect to the present invention by their mention in this background.)

Patents That May Be Helpful In Starting A Search Of The Background

Some patents that are in the same general areas as versions of the invention are cited here: US Patent Number 5,667,976 Solid supports for nucleic acid hybridization assays. Published International Application WO 98/20165 Biallelic Markers. Published International Application WO 98/07887 Methods for treating bipolar mood disorder associated with markers on chromosome 18 p. US Patent Number 5,552,270 Methods of DNA sequencing by hybridization based on optimizing concentration of matrix-bound oligonucleotide and device for carrying out same. No patent in this paragraph is admitted to being prior art with respect to the present invention by it's mention in this background.

Summary

Versions of the invention use a new, two-dimensional concept of "closeness" for association-based linkage studies. Versions of the invention use bi-allelic markers that "cover" or are distributed approximately evenly (or systematically) over two-dimensional regions. These regions have the two dimensions of chromosomal location and least common allele frequency.

Conventional techniques suffer from a kind of one-dimensional lack of depth perception. (They also favor bi-allelic markers with least common allele frequencies near 0.5.) Two-dimensional linkage study techniques overcome this lack of depth perception. These two-dimensional techniques greatly increase the chance that one or more markers used in a study will be close to the sought gene in two-dimensions. This results in more powerful, systematic and efficient methods (including computer programs) and machines for finding genes, such as harmful genes and genes of only modest effect. These techniques also use less dense (more efficient) marker maps (or marker "coverings").

The basic principles behind the two-dimensional approach spawn numerous other inventions. These include methods, machines and compositions of matter (groups of molecules) used for gathering the data (i.e. genotype/sample allele frequency data) used in the new two-dimensional studies, and computer techniques for using and handling such data. These techniques work for creatures other than human beings. And they work for markers and genes that are not bi-allelic (any marker or gene can be mathematically transformed to behave like it is bi-allelic). This summary is not exhaustive or limiting, there are other inventions not listed or specifically described here.

---

[12] Falk CT and Rubenstein P: Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. Annals of Human Genetics, 1987, vol. 51, pp. 227-233.

[13] Bell GI, Horita S and Karam JH: A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. Diabetes, 1984, vol 33, pp. 176-183.

A CL-F matrix is a matrix of rectangular cells of the same length and the same width on a CL-F map. Stipulations that a certain number of covering markers are placed in each cell of the matrix is a method of illustrating particular types systematic covering of a CL-F region with covering mark rs.

The evidence for linkage obtained from two-dimensional linkage studies is essentially two-dimensional in nature and it is possible to use this two-dimensional information by essentially graphing quantitative evidence for linkage as a function of position in the x-y plane. For example, if quantitative evidence for linkage is represented in the z dimension of a typical three-dimensional x-y-z plot, wherein the x and y dimensions are chromosomal location and least common allele frequency respectively, then it is possible to conceptualize evidence for linkage as occurring in a "hump" or "humps" in the z dimension. And it is possible to analyze the data to find the CL-F location (in the x-y plane) of the peak(s) of this "hump(s)", thus helping to localize a trait causing gene to the CL-F locale of the peak(s) of the "hump(s)".

Versions of the invention also make use of multi-allelic genes and/or markers. It is always possible to combine the alleles of a multi-allelic polymorphism (marker or gene) so that the polymorphism acts mathematically like it is a bi-allelic polymorphism. In effect, it is always possible to mathematically transform a multi-allelic marker or gene to act bi-allelic. Similarly, two or more markers can always be mathematically combined to form a mathematical marker that acts like a single bi-allelic marker. And two or more genes can always be mathematically combined to form a mathematical gene that acts like a single bi-allelic gene. In this application a mathematical bi-allelic marker formed mathematically from one or more markers is called a bi-allelic marker equivalent or BME; and a mathematical bi-allelic gene formed mathematically from one or more genes is called a bi-allelic gene equivalent or BGE.

The term true marker or gene is used to distinguish a marker or gene in the ordinary sense from a bi-allelic marker equivalent (BME) or bi-allelic gene equivalent (BGE). The term true allele is used to distinguish an allele in the ordinary sense from a mathematical allele of a BME or BGE. A mathematical allele of a BME or BGE is referred to as an allele equivalent. An allele equivalent is a combination of one or more true alleles or one or more haplotypes.

Versions of the invention make use of genes and/or markers, which are not exactly bi-allelic. These genes or markers are approximately bi-allelic. A gene or marker that is approximately bi-allelic almost always occurs in one of two allele forms, however, very rarely it occurs in a different allele form.

Various versions of the invention are for genotyping individuals at markers which systematically cover CL-F regions or for obtaining sample allele frequency data (such as from pooled DNA) for a sample of individuals for markers which systematically cover CL-F regions. Various versions of the invention are for oligonucleotides used for genotyping individuals at markers which systematically cover CL-F regions or are for obtaining sample allele frequency data (such as from pooled DNA) for a sample of individuals for markers which systematically cover CL-F regions.

D finitions

For the purposes of the description and claims the terms used herein will have their generally accepted definition unless otherwise specified.

If **a CL-F region is said to comprise an area of greater than or equal to X multiplied by Y**, then the CL-F region comprises one or more nonoverlapping segment-subranges, and the sum of the areas of the segment-subranges is greater than or equal to X multiplied by Y.

A **CL-F matrix** is a collection of segment-subranges, wherein each segment-subrange of the collection has the same width and the same length. Each segment-subrange in the collection (or the matrix) is a **CL-F matrix cell.** Any one CL-F matrix cell in a CL-F matrix shares two or more of the cell's borders with two or more other cells in the matrix. And all the cells in a CL-F matrix together form a single segment-subrange. A CL-F matrix is characterized by the length and the width of the cells in the denoted by length x width, or $L_{MC}xW_{MC}$, wherein L_{MC} is the length of each cell in the matrix and W_{MC} is the width of each cell in the matrix. A CL-F matrix is also characterized by the number of rows of cells, R_M, in the matrix. And a CL-F matrix is characterized by the number of columns of cells, C_M, in the matrix. There are two or more cells in a CL-F matrix. A CL-F matrix is also characterized by the point of origin of the matrix, denoted by (cl_O, f_O). The point of origin of a CL-F matrix is at any chromosomal location and cl_O takes on any reasonable value in an entire species genome. The point of origin of a CL-F matrix is at any one value in the least common allele frequency range 0 to 0.5. (A CL-F matrix is similar to the squares of a chessboard or to equal rectangular floor tiles that are all oriented in the same direction and cover a rectangular floor. One corner of the matrix is the matrix's point of origin.)

The **width of each cell of a particular CL-F matrix** is any value greater than zero and less than 0.5. The width of a cell is often denoted by W_{MC}.

Any length in chromosomal location distance units is chosen for the **length of each cell of a particular CL-F matrix**. The length of a cell is often denoted by L_{MC}.

The **centerpoint** of a CL-F matrix cell is in the center of the cell. The centerpoints of a CL-F matrix form a **matrix centerpoint lattice.** Each point of a matrix centerpoint lattice is separated by a CL-F distance of [0, W_{MC}] or [L_{MC}, 0] from two or more neighboring centerpoints.

If **one or more bi-allelic markers are in(or within) the segment-subrange that is a CL-F matrix cell,** then each of the markers is in or within the CL-F matrix cell.

If **one or more CL-F points is in (or within) a CL-F matrix**, then each of the points is in or within a cell of the matrix.

If **a CL-F region comprises a CL-F matrix**, then each point that is in the matrix is also in the region.

If **a CL-F region is a CL-F matrix**, then the region consists of the points that are in the matrix.

If two CL-F matrix cells share a common border, then the **two CL-F matrix cells are in contact.**

If two CL-F matrix cells share a common corner, then the **two CL-F matrix cells are touching.** (Two cells that are in contact are also touching.)

If **a group of CL-F points is connected to within a CL-F distance [X,Y]**, then for any two points in the group, denoted p_1 and p_R, there is an ordered sequence of points in the group denoted p_1, p_2, p_3,....., p_{R-2}, p_{R-1}, p_R, R being an integer greater than or equal to 2, wherein the CL-F distance between each point in the sequence and the next point in the sequence is less than or equal to [X,Y]. **The distance [X,Y] is the connecting distance.** (Put in simple terms if a group of points is connected to within [X,Y], then there is a path between each pair of points in the group, the path consisting of a series of steps, wherein each step in the path is a movement between two points in the group that are

separated by a CL-F distance of less than or equal to [X,Y]. A simple group of points connected to within a CL-F distance of [X,Y] is a group of three points, wherein each point in the group is within a CL-F distance of less than or equal to [X,Y] of another point in the group. The concept of connectivity introduced here is similar to the basic concept of connectivity in mathematical graph theory.)

If **a group of N markers is connected to within a CL-F distance [X,Y]**, wherein N is an integer, then each of the markers is located at one point of group of N points, the group of N points being connected to within a CL-F distance [X,Y].

If **two bi-allelic markers are said to be in extreme positive disequilibrium** then d is approximately equal to d_{max} for the two markers, which for the purposes of this definition are designated marker M with least common allele A and marker m with least common allele B. Wherein according to standard usage, the disequilibrium coefficient (d) is defined by the equation d=f(AB) - f(A)f(B) where f(A) and f(B) are defined as the population frequencies of alleles A and B, respectively, and f(AB) is the population frequency of the AB haplotype. And d_{max} is defined as the maximum possible positive value of d assuming the allele frequencies of A and B are f(A) and f(B), and thus d_{max}= q-f(A)f(B) where q is the lesser of f(A) and f(B). (In this application d is used to represent the disequilibrium coefficient; the symbol δ is often used in scientific papers to represent the disequilibrium coefficient.)

If **a pair of markers is said to be in extreme positive disequilibrium**, then the two markers of the pair are in extreme positive disequilibrium.

If **a pair of bi-allelic markers is said to be redundant within distance D** then the two markers of the pair are in extreme positive disequilibrium and the two markers are located on the same chromosome and the two markers are located within a CL-F distance D of each other on a CL-F map, wherein D is a specified distance and D has two components, a chromosomal location distance component D_{CL} and a frequency distance component, D_F; D = [D_{CL}, D_F].

An **allele equivalent (AE)** is a group of one or more "haplotype values" of one or more polymorphisms of the same type, either markers or genes. (For the purposes of this application a haplotype value of one polymorphism is equivalent to an allele value at the one polymorphism.)The group of haplotype values is then analyzed as if the group is a single allele at a bi-allelic polymorphism; the group of haplotype values acts as a single allele at a bi-allelic polymorphism; the collection of the one or more polymorphisms upon which the haplotype values are based acts as a bi-allelic polymorphism; the collection of one or more polymorphisms forms a bi-allelic **polymorphism equivalent (PE)** that acts as a bi-allelic polymorphism; **the polymorphism equivalent has(or possesses) the allele equivalent.** The allele equivalent belongs to the polymorphism equivalent. In this application, **each polymorphism equivalent is a bi-allelic marker equivalent(BME) or a bi-allelic gene equivalent(BGE).**

A bi-allelic marker equivalent (BME) is one or more markers and a grouping of the haplotype values of the one or more markers into two groups (e.g. group I and group II)(For the purposes of this application a "haplotype value" of one marker is equivalent to an allele at the one marker). The one or more markers and the two groups of haplotype values of the on or more markers are then analyzed as if the one or more markers are a single bi-allelic marker with alleles I and II. Each group of the groups I and II is an allele equivalent. For example, a multi-allelic microsatellite marker has it's multiple alleles grouped into two groups and the microsatellite marker and these two groups of alleles then act

details regarding this, see Detailed Description of the Systematic Covering of a CL-F Region Used In Versions of the Invention above.

An example of ProcessGd/Safd#1Genotype data/Sample allele frequency data process, a genotype data process:

Example 1 of ProcessGd/Safd#1:A process for obtaining genotype data/sample allele frequency data for each bi-allelic marker of a group of two or more bi-allelic covering markers in the chromosomal DNA of an individual, wherein the genotype data/sample allele frequency data is genotype data, comprising:

a) means for determining information on the presence or absence of each allele of each bi-allelic marker of a group of two or more bi-allelic covering markers in the chromosomal DNA from an individual, a CL-F region being N covered to within the CL-F distance [12cM, 0.25] or the equivalent thereof by the two or more bi-allelic covering markers; wherein N is an integer number greater than or equal to 1; and

b) means for transforming the information of step a) into genotype data for each marker of the group.

(It should be noted that the following genotype process is equivalent to Example 1 of ProcessGd/Safd#1: Genotype Process: A process for genotyping an individual, comprising: a)means to genotype an individual at two or more bi-allelic covering markers, a CL-F region being N covered to within the CL-F distance [12cM, 0.25] or the equivalent thereof by the two or more bi-all lic covering markers, wherein N is an integer number greater than or equal to 1.)

Oligonucleotide technology

Each version of oligonucleotide technology is a means to sense the presence or absence of each of one or more true alleles of a group of true alleles in chromosomal DNA from one or more individuals by means of a hybridization reaction with an oligonucleotide that is complementary to each of the one or more true alleles (see definitions section). Thus versions of oligonucleotide technology are a means of genotyping one or more individuals. And, versions of oligonucleotide technology are a means of obtaining sample allele frequency data for one or more marker alleles for a sample of individuals using pooled DNA from the individuals in the sample.

In Some Versions of Oligonucleotide Technology for Genotyping or Obtaining Sample Allele Frequency Data, a Physico-chemical Signal is Generated when an Allele in Chromosomal DNA and a Complementary Oligonucleotide Hybridize

Some versions of oligonucleotide technology for genotyping or for obtaining sample allele frequency data use a sensor which includes one or more oligonucleotides which are complementary to an allele. When the s nsor is exposed to chromosomal DNA from an individual who carri s the allele, the oligonucl otid s which are compl m ntary to th all le hybridiz with chromosomal DNA specimens of the all le. The hybridization generates a physico-chemical signal which indicates the presence of the

allele in the chromosomal DNA of the individual. The lack of the physico-chemical signal indicates no (or negligible)hybridization and that the allele is not present in the chromosomal DNA of an individual.

Examples of oligonucleotide technology for genotyping, obtaining sample allele frequency data or genotype data/sample allele frequency data

Companies like Affymetrix are using high density arrays of oligonucleotides attached to silicon chips or glass slides to genotype DNA from one individual at thousands of bi-allelic markers.[VIII] In some of these versions of oligonucleotide technology, the strength of hybridization of oligonucleotides that differ at only one base to DNA containing an SNPare compared to determine genotype.[IX] Another version of oligonucleotide technology uses oligonucleotides as PCR (Polymerase Chain Reaction) primers to obtain genotype data.[X]Other examples of oligonucleotide technology and it's uses to obtain genetic information are included in the articles cited in the endnotes.[XI] Versions of oligonucleotide technology obtain sample allele frequency data from pooled DNA or genotype data using oligonucleotides as PCR primers to obtain amplified reaction products that are detected by mass spectrometry. Another example of oligonucleotide technology is padlock probes.[XII]

Other examples of oligonucleotide technology are minisequencing on DNA arrays, dynamic allele-specific hydridization, microplate array diagonal gel electrophoresis, pyrosequencing, oligonucleotide-specific ligation, the TaqMan system and immobilized padlock probes as presented at the First International Meeting on Single Nucleotide Polymorphism and Complex Genome Analysis.[XIII]

Sets of Oligonucleotides for Genotyping at Bi-allelic Markers or Obtaining Sample Allele Frequency Data

A *set of oligonucleotides that is complementary(see definitions) to a group of one or more bi-allelic markers has utility to determine genotype data at each of the markers in the group*, including groups with BMEs and approximately bi-allelic markers.

Similarly, a set of oligonucleotides that is complementary to a group of bi-allelic markers has utility to obtain sample allele frequency data for each allele of each marker in the group.

In both cases, obtaining genotype data or sample allele frequency data, the same principle is used: a set of oligonucleotides that is complementary to a group of bi-allelic markers has utility determine the presence or absence of each allele of each marker in the group in chromosomal DNA.

Using sets of oligonucleotides to obtain Genotype Data/Sample Allele Frequency Data for each marker of a group of bi-allelic markers, wherein the group of markers systematically cover a CL-F region

Genotype data/sample allele frequency data for each marker of a group of bi-allelic markers, wherein the group of bi-allelic markers systemically cover a CL-F region has great utility for use in the more powerful two-dimensional linkage studies introduced in this application. As described above under Oligonucleotide Technology, some sets of oligonucleotides have utility to determine genotype data at each bi-allelic marker of a group of one or more bi-allelic markers. Similarly, some sets of oligonucleotid s have utility to obtain sample allel frequency data for each bi-allelic marker of a group of one or more bi-allelic markers. Therefore, the use of one or more copies of a set of oligonucleotides to obtain genotype data or sample allele frequency data for each bi-allelic marker of a group of one or

Versions of the apparatus comprise means for printing each of the one or more graphs.

The ry of Operation / Set/Subset Example

Systematically Varying Both Marker Chromosomal Location and Marker Allele Frequency of Markers in Linkage Studies

The inventor's calculations and observations have demonstrated the increased power of the TDT in more common, less optimal situations when a bi-allelic marker and bi-allelic gene have (1) similar but not identical allele frequencies and (2) the marker and gene are in some degree of linkage disequilibrium. Thus, for a typical linkage study using bi- allelic markers and an association based linkage test, ***to increase the likelihood of both criteria (1) and (2) occurring for one or more markers, so as to increase the power of an association based linkage test in a linkage study, the bi-allelic markers used in the study are chosen so that the least common allele frequencies of the markers vary systematically over a range or subrange of least common allele frequency AND the chromosomal location of the markers vary systematically over one or more chromosomes or chromosomal regions. And the bi-allelic markers are chosen so that the markers' chromosomal locations and least common allele frequencies vary systematically in an essentially independent manner.***

(In theTheory of Operation/ Set/Subset Example Section the traditional symbol used in scientific papers for the disequilibrium coefficient, δ, is used. This should not be confused with the symbol δ used for the covering distance in the remainder of the application. The symbol d is used for the disequilibrium coefficient in the sections of the application other than the Theory of Operation/Set/Subset Example Section.) The theory of operation is based on the mathematical observation that the TDT and other association-based tests for linkage are increased in power as the frequencies of the disease-causing allele of a bi-allelic gene and the positively associated allele of a linked bi-allelic marker become similar in magnitude. The inventor made this observation as a result of deriving the equation shown below for P_t (this is Equation 2 in the unpublished manuscript submitted for publication in December 1996 and in published paper by RE McGinnis in the Annals of Human Genetics vol 62, pp. 159-179, 1998).

$$P_t = .5 + (1 - 2\theta)\left[\frac{c_1c_4 - c_2c_3}{H}\right]\left\{p^2\left(\frac{\alpha^2 - \beta^2}{4}\right) + 2p(1-p)\left(\frac{(\alpha+\beta)^2 - (\beta+\gamma)^2}{16}\right) + (1-p)^2\left(\frac{\beta^2 - \gamma^2}{4}\right)\right\}$$

Equation 2

P_t may be regarded as the size of the "signal" which is given by the TDT to indicate that a tested marker is linked to a disease-causing gene. The more P_t is elevated above 0.5 (baseline), the greater is the evidence for linkage or "power" provided by the association-based linkage test known as the TDT.

Table 2 in the unpublished manuscript filed with previous US Provisional Patent Applications(see below) illustrates how signal strength increases substantially as the frequencies of disease-causing allele and positively associated marker allele become similar in magnitude. As noted on pages 24 and 25 of the unpublished manuscript(see below), Table 2 assumes that the frequency (p)

of the disease-causing allele is fixed at p=.1 while the frequency (m) of the positively associated marker allele varies (m=.5, .3, .2, .1, .05). Note that when the level of disequilibrium (or association) between the bi-allelic marker and bi-allelic disease gene is fixed (in this case either $\delta=\delta_{max}$ or $\delta=\frac{1}{2}\delta_{max}$), the signal strength of P_t progressively increases as m decreases from m=.5 to m=.1 (the same frequency as the disease allele, i.e., p=.1). For example, in the section of Table 2 for r=5, note that when $\delta=\frac{1}{2}\delta_{max}$, P_t is .548 at m=.5 and then steadily increases to .572 (m=3), .597 (m=.2), .648 (m=.1) and then starts to decrease again as m departs from m=p=.1 (i.e. P_t =.636 at m=.05). As noted on pages 24-25 (below)of the unpublished manuscript, the TDT chi-square statistic (assuming a sample size of 200 families) is such that the signal strength at m=.5 (P_t =.548) does not produce a statistically significant evidence for linkage (p-value > 0.5) while the doubling of signal strength at m=.2 (P_t =.597) produces very strong statistical evidence for linkage by the TDT (p-value< 0.005). This sort of substantial increase in power is also true of other association-based linkage tests as the frequencies of the disease-causing allele and associated marker allele become more similar in magnitude.

judicious use of bi-allelic markers of both high, medium, and low heterozygosity may be crucial in order to initially detect and replicate linkages to loci conferring modest disease risk.

Set/Subset Example:

Method for locating disease causing polymorphism using biallelic linkage analysis

Objective :To test, by association-based linkage analysis (e.g., by TDT), whether a disease-causing polymorphism is located on a particular chromosome (e.g., human chromosome 4) or within a particular subregion of that chromosome.

PART 1 - Steps in conducting the association-based linkage test

Step 1

To conduct the test, first divide the chromosome or subregion of interest into segments that are short enough that polymorphisms within each segment are likely to be in linkage disequilibrium with each other. The division of a chromosome or subregion of interest into "segments" is conceptual (*not* physical) and is based on chromosomal maps such as those provided by the Whitehead Institute or Marshfield Foundation for Biomedical Research. Although disequilibrium has been observed in Finnish populations between polymorphisms that are 7 to 10 centimorgans (cM) apart, the chromosomal segments for searching for disease-causing polymorphisms in more genetically heterogeneous populations should be less than 1 cM long (e.g., 250,000 base pairs long). These chromosomal segments might or might not overlap each other (i.e., share some of their length in common); but the set of chromosomal segments should completely cover the entire chromosome or entire subregion of interest, so that a disease-causing polymorphism located anywhere on the chromosome or anywhere in the subregion of interest will be detected by the test.

Step 2

It is well known that increased disequilibrium between a marker and linked disease locus increases evidence for linkage provided by association-based linkage tests such as the TDT. However, what has not been recognized is that the specific allele frequencies of the marker locus can also have an enormous impact on the strength of evidence for linkage. I

the nearly identical information with respect to their linkage and association with a third polymorphism such as a disease locus. Hence one of the two bi-allelic markers would provide no additional information and its inclusion in the subset would not increase the likelihood of detecting linkage and association to a nearby disease locus.

Therefore, bi-allelic markers belonging to the same chromosomal segment and subset should not only have similar allele frequencies, the δ value between *each pair* of bi-allelic markers in the same subset should be substantially less than $\delta_{max}= q-q^2$. This assures that every bi-allelic polymorphism belonging to the subset provides much new (i.e. non-redundant) information about linkage and association to any nearby bi-allelic disease locus; thus testing each bi-allelic marker in the subset would increase the likelihood of detecting linkage to a disease locus.

Step4: Test for linkage

To test for (association-based) linkage to a bi-allelic disease locus, each bi-allelic marker in each subset from each chromosomal segment is tested *individually* by using the TDT, AFBAC method or other family-based linkage test. To conduct these tests for a particular marker, members of nuclear families (most especially parents, and any children who manifest disease) are genotyped at the marker being tested and the genotypes are then evaluated according to the TDT, AFBAC method or other family-based linkage/association test (for description of TDT and AFBAC, see Spielman et al, Am J of Human Genetics 52:506-516 (1993) and Thomson, Am J Human Genetics 57:487-498 (1995)). Alternatively, linkage and association is tested for each marker in each subset from each segment by genotyping individuals with disease and related or unrelated normal controls at each marker to be tested. (End of set/subset example)

Further Information

(Step 3 is not essential for the operation or utility of this version of the invention. In this set/subset example, the least common allele frequency subrange 0.1 to 0.5 is used. In versions of the invention similar to the set/subset example, versions of the invention are operable and have utility for any subrange of the least common allele frequency range 0 to 0.5. In addition, rather than genotyping DNA from single individuals in step 4, in some versions of the invention each marker in each subset from each segment is tested for association with disease by evaluating DNA from pooled samples.)

Statement under Article 19(1)

PCT/US99/04376

Some of the amended claims make use of the phrase "conditional probability", such as claim 11. Some of the amended claims make use of the phrase "proportion of groups", such as claim 14. There are various techniques to calculate or estimate such a probability or such a proportion. These techniques include, but are not necessarily limited to, direct calculation, statistical estimates, and Monte Carlo estimation techniques. Powerful software is available for calculation and statistical estimation for data in matrix format or two-dimensional format. Some such software is available from Cytel Software Corporation, Cambridge, Massachusetts (example: Exact Logistic Regression: Theory and Examples, Mehta CR, Patel NR, Statistics in Medicine, vol 14, 2143-2160(1995). Another example is SAS (SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513, USA.; A handbook of statistical analyses using SAS by Brian S. Everitt and Geoff Der, Boca Raton, Fla. : Chapman & Hall/CRC, 1998.). A further example is MATLAB (The MathWorks, Inc. 3 Apple Hill Drive, Natick, Mass. U.S.A. 01760-2098; MATLAB primer by Kermit Sigmon, 4th ed. Boca Raton : CRC Press, c1994.) Statistical techniques include techniques for hypothesis testing, goodness-of-fit and others.

The degree of skill in the art in probability and statistics is great. Indeed the inventor's important equation (Equation 2, page 38) is an equation for P_t, wherein P_t is a binomial probability for parental allele 'transmission' which determines the magnitude of the TDT chi-square statistic. P_S (pages 40-42) is also a binomial probability that determines the magnitude of the ASP test statistic. (see Abstract and Paper: Annals of Human Genetics (1998), 62, 159-179. The abstract is available on the World Wide Web and Internet, including at the journal's website.) Skill in the use of computers in the art is also great (page 25).

Some claims, such as claims 11, 12, 13, 14 and others make use of the phrase "substantially the known set of bi-allelic markers". As pointed out in the description (page 25) information on bi-allelic markers can be gained from sources such as the Whitehead Institute or Marshfield Foundation for Biomedical Research. Similar sources of information on Single Nucleotide Polymorphisms can be obtained from sources given in SNP attack on complex traits, Nature Genetics, volume 20 no. 3, Nov 1998, pp. 217-218.

Some claims, such as claims 11, 12, 13, 14 and others make use of the term "marker type" or similar terminology. As stated in the description, a bi-allelic marker may be an SNP, a microsatellite marker, a bi-allelic marker equivalent formed from one or more true bi-alleic markers. "Marker type" means type of true bi-allelic marker as for example an SNP or a microsatellite; or "marker type" means a bi-allelic marker equivalent of a certain type, such as a bi-allelic marker equivalent formed only from one or more SNPs or a bi-allelic marker equivalent formed only from one or more microsatellites.)